Human Activity Recognition: A Review

Amr Sharaf
Computer and Systems Engineering Department
Faculty of Engineering
Alexandria University
Email: amr.nabil.sharaf@gmail.com

***Abstract*—Human activity recognition is an important research area in the field of computer vision. Applications of human activity recognition include but are not limited to surveillance systems, patient monitoring and human machine interaction. This report provides an overview of various research papers on human activity recognition. The surveys by Gavrila, Aggarwal et al. and by Moeslund et al. provide a broad overview for hundreds of papers and numerous approaches for analyzing human activity in videos, including human motion. This report summarizes the most important methodologies used for activity recognition based on these surveys.**

I. Introduction

Human activity recognition is an important research area in the field of pattern recognition and computer vision. The goal of human activity recognition is to automatically analyze ongoing activities from an unknown video. In simple case where a video is segmented to contain only one execution of a human activity, the objective of the system is to correctly classify the video into activity category. In more general cases, the continuous recognition of human activities must be performed, detecting starting and ending times of all occurring activities from an input video.

II. Applications

The ability to recognize complex human activities from videos enables the construction of several important applications. Automated surveillance systems in public places like airports and subway stations require detection of abnormal and suspicious activities as opposed to normal activities. For instance, an airport surveillance system must be able to automatically recognize suspicious activities like 'a person leaving a bag' or 'a person placing his/her bag in a trash bin'. Recognition of human activities also enables the real-time monitoring of patients, children and elderly persons. The construction of gesture-based human computer interfaces and vision-based intelligent environments becomes possible as well with an activity recognition system.

III. Types of Human Activities

There are various types of human activities. Depending on their complexity, human activities could be categorized into four different levels: gestures, actions, interactions and group activities. Gestures are elementary movements of person's body part, and are the atomic components describing the meaningful motion of a person. 'Stretching an arm' and 'raising a leg' are good examples of gestures. Actions are simple person activities that may be composed of multiple gestrues organized temporally, such as 'walking', 'waving' and 'punching'. Interactions are human activities that involve two or more persons and/or objects. For example, 'two persons fighting' is an interaction between two humans and 'a person stealing a suitcase from another' is a human-object interaction involving two humans and one object. Finally, group activities are the activities performed by conceptual groups composed of multiple persons and/or objects. 'A group of persons marching', 'a group having a meeting', and 'two groups fighting are typical examples of them.

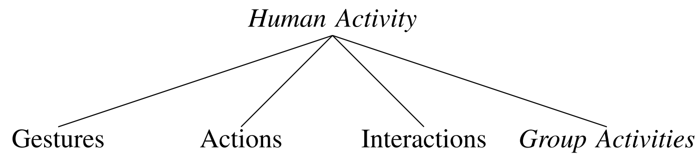

Fig. 1. Different types of human activities.

IV. Approach-based Taxonomy for Human Activity Recognition

Figure 2 illustrates an overview of the tree-structured taxonomy that J. K. Aggrawal and M. S. Ryoo review follows. They have chosen an approach based taxonomy. All activity recognition methodologies are first classified into two categories: single-layered approaches and hierarchical approaches. Single-layered approaches are approaches that represent and recognize human activities directly based on sequences of images. Due to their nature, single-layered approaches are suitable for the recognition of gestures and actions with sequential characteristics. On the other hand, hierarchical approaches represent high-level human activities by describing them in terms of other simpler activities, which they generally call sub-events. Recognition systems composed of multiple layers are constructed, making them suitable for the analysis of complex activities.

Single-layered approaches are again classified into two types depending on how they model human activities: space-time approaches and sequential approaches. Space-time approaches view an input video as a 3-dimensional (XYT) volume while sequential approaches interpret it as a sequence of observations. Space-time approaches are further divided into three categories based on what features they use from the 3-D space-time volumes: volumes themselves, trajectories, or local interest point descriptors. Sequential approaches are classified depending on whether they use exemplar-based recognition methodologies or model-based recognition methodologies.

Hierarchical approaches are classified based on the recognition methodologies they use: statistical approaches, syntac-

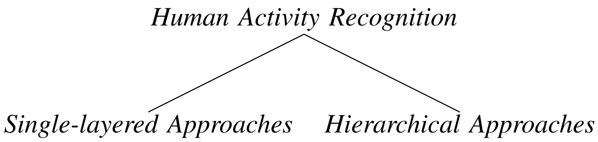

Fig. 2. Different types of human activities.

tic approaches, and description-based approaches. Statistical approaches construct statistical state-based models concatenated hierarchically (e.g. layered hidden Markov models) to represent and recognize high-level human activities. Similarly, syntactic approaches use a grammar syntax such as stochastic context-free grammar (SCFG) to model sequential activities. Essentially, they are modeling a high-level activity as a string of atomic-level activities. Description-based approaches represent human activities by describing sub-events of the activities and their temporal, spatial, and logical structures.

V. Comparison

In general, sequential approaches consider sequential relationships among features in contrast to most of the space-time approaches, thereby enabling detection of more complex activities (i.e. non-periodic activities such as sign languages). Particularly, the recognition of the interactions of two persons, whose sequential structure is important, has been attempted in [Oliver et al. 2000; Natarajan and Nevatia 2007].

Compared to the state model-based sequential approaches, exemplar-based approaches provide more exibility for the recognition system, in the sense that multiple sample sequences (which may be completely different) can be maintained by the system. Further, the dynamic time warping algorithm generally used for the exemplar-based approaches provides a non-linear matching methodology considering execution rate variations. In addition, exemplar-based approaches are able to cope with less training data than the state model-based approaches.

On the other hand, state-based approaches are able to make a probabilistic analysis on the activity. A state-based approach calculates a posterior probability of an activity occurring, enabling it to be easily incorporated with other decisions. One of the limitations of the state-based approaches is that they tend to require a large amount of training videos, as the activity they want to recognize gets more complex.

Hierarchical approaches are suitable for recognizing high-level activities which can be decomposed into simpler sub-events. Because of their nature, they can more easily incorporate human knowledge into the systems and require less training data as pointed out by many researchers. Statistical and syntactic approaches provide a probabilistic framework for reliable recognition with noisy inputs. However, they have difficulties representing and recognizing activities with concurrently organized sub-events.

Description-based approaches are able to represent and recognize human activities with complex temporal structures. Not only sequentially occurring, but also concurrent organized sub-events are handled with description-based approaches. The major drawback of description-based approaches are their inability to compensate for the failures of low-level components (e.g. gesture detection failure). That is, most of the description-based approaches have a deterministic high-level component. Pinhanez and Bobick [1998] showed that the high-level system has the potential to compensate for a single low-level detection failure, and a couple of recent works have proposed probabilistic frameworks for description-based approaches [Ryoo and Aggarwal 2009a; Gupta et al. 2009].

VI. Case Study: Human Activity Recognition Using Histogram of Optical Flow

In this case study we will perform some basic action recognition using motion features.

A. Dataset

A portion of the Weizmann Human Action Classification Dataset has been used. 3 instances of 3 different actions (run, walk, jump) have been selected.

B. Human Detection Via Background Subtraction

A simple human detector has been implemented by computing the median of the foreground for each frame of each video. A simple method that is ective for this dataset is to take the median value of each pixel of a video over time as a background image, then threshold the absolute value of differences to this background image.

C. Histograms of Optical Flow

The code in "lk3.m" has been used to compute the optical flow on a video. The paper "Histograms of Oriented Optical Flow and Binet-Cauchy Kernels on Nonlinear Dynamical Systems for the Recognition of Human Actions" suggeted using the histogram of optical flow to avoid scale and direction of motion invariance. The code "gradiantHistogram.m" has been used to calculate the histogram of optical flow.

D. Classification

Per-frame classification has been performed of the video frames into the 3 different categories. Nearest neighbour classifier has been used to classifiy each frame. Two videos per category have been used for training the classifier and the frames from the remaining third video have been used for testing.

E. Results

Prediction outcome

		p	**n**	**total**
actual value	**p′**	31	70	P′
	n′	14	8	N′
	total	P	N	

Precision	0.6889
Recall	0.3069
Correct Rate	0.6098
Error Rate	0.3902

VII. Conclusion

Computer recognition of human activities is an important area of research in computer vision with applications in many diverse fields. The application to surveillance is natural in today's environment where the tracking and monitoring people is becoming an integral part of everyday activities. Other applications include human-computer interaction and biometrics based on gait or face, and hand and face gesture recognition.

A significant amount of progress on human activity recognition has been made in the past 10 years, but it is still far from being an off the shelf technology. We are at a stage where experimental systems are deployed at airports and other public places. It is likely that more and more, such systems will be deployed.

The future direction of research is encouraged and dictated by applications. The pressing applications are the surveillance and monitoring of public facilities like train stations, underground subways or airports, monitoring patients in a hospital environment or other health care facilities, monitoring activities in the context of UAV surveillance, and other similar applications. All of these applications are trying to understand the activities of an individual or the activities of a crowd as a whole and as subgroups.